

データマネジャーではない人へ

試験準備段階のCDM業務 1

-CDM業務の全体像-

-データベースの理解(基本編)-

国立がん研究センター 中央病院

臨床研究支援部門

データ管理部

加幡晴美

2019/5/10

内容

1. CDM (Clinical Data Management) の全体像
2. データベースの基本
3. CRF設計
 - CRF の要件
 - CRF 設計の手順
 - 紙 CRF
 - EDC (Electronic Data Capture)

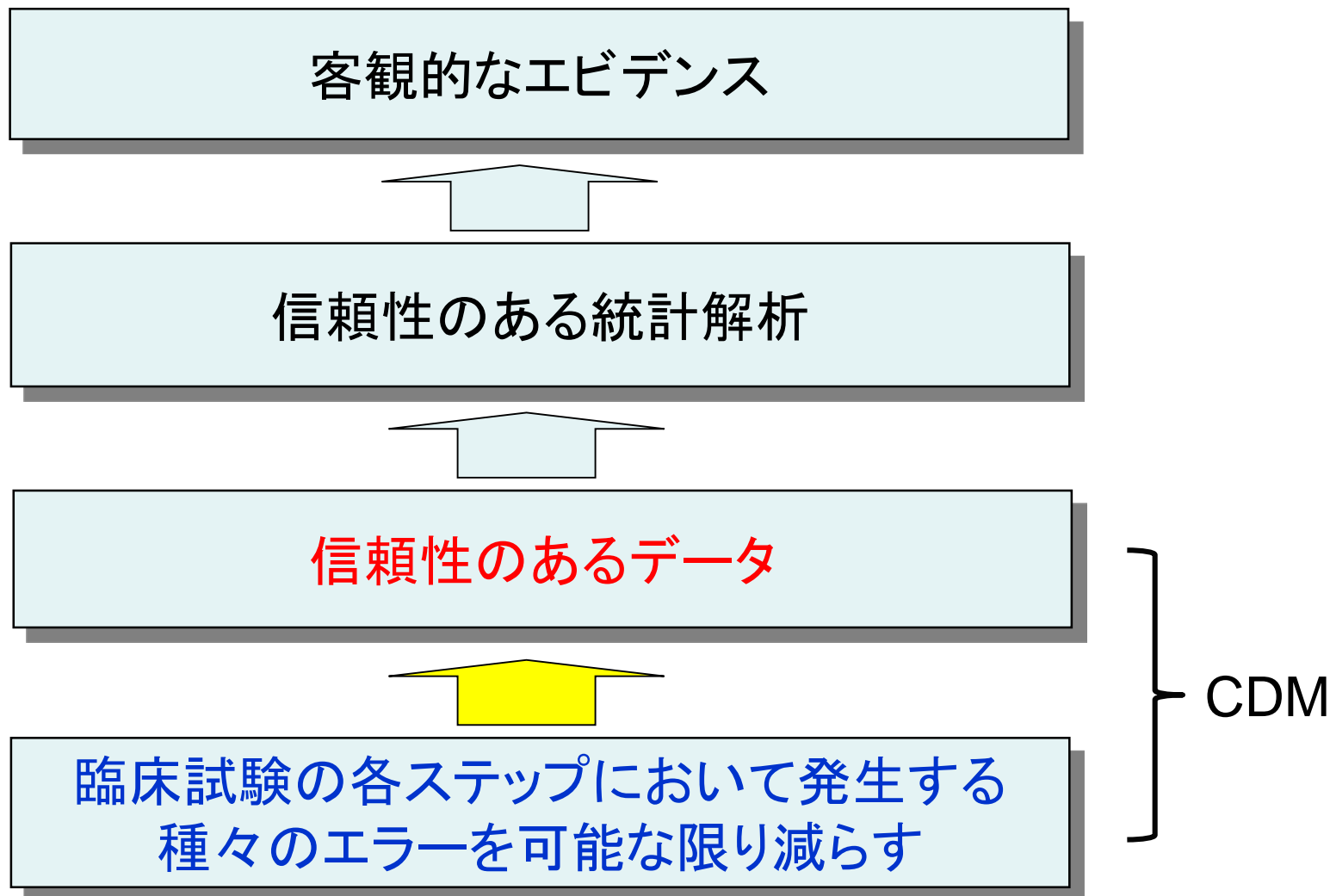
CDM はデータの品質管理の中心・司令塔

- でも、CDM だけでデータの品質を管理できる訳ではない
- 試験に関わる担当者がそれぞれの役割（責務）をしっかりと果たすことが大切！
- 研究者（治験責任医師、治験担当医師）、CRC、モニター、統計担当者など

臨床試験で求められる品質のレベル

- 被験者の人権の保護
- 安全性の確保
- 臨床試験の科学的な質、
つまり成績の信頼性の確保

臨床試験の目的と CDM の役割



臨床試験における CDM の具体的な業務

- ✓ プロトコールのレビュー
- ✓ CRFの見本の設計
- ✓ CRF記載マニュアルの作成
- ✓ DB(EDC)の構築
- ✓ データ入力ルール作成
- ✓ データチェック規定の策定と実施
 - チェック項目の設定
 - チェックの実施、クエリーの発行
- ✓ 中央モニタリングの実施・支援
- ✓ コーディング、辞書管理
- ✓ データ固定

プロトコールレビュー

- 論理的整合性
 - 実施可能性
 - プロトコールと CRF との整合性
必要十分なデータを適切に入手できるのか？
-
- ✓ 選択規準、除外規準
 - ✓ 登録・割付
 - ✓ 治療計画と治療変更規準
 - ✓ 評価項目・臨床検査・評価スケジュール
 - ✓ 効果判定とエンドポイントの定義

内容

1. CDM (Clinical Data Management) の全体像
2. データベースの基本
3. CRF設計
 - CRF の要件
 - CRF 設計の手順
 - 紙 CRF
 - EDC (Electronic Data Capture)

データベースとは？

- データ



- ベース

- データベースがあると何ができる？

非専門家にありがちな例

登録番号	イニシャル	性別	生年月日	既往歴
1	SH	男性	1970/2/5	喘息
2	YA	女性	1955/7/3*	高血圧
3	KO	男	1945/12/1	糖尿病

非専門家にありがちな例

登録番号	イニシャル	性別	生年月日	既往歴
1	SH	男性	1970/2/5	喘息
2	YA	女性	1955/7/3*	高血圧
3	KO	男	1945/12/1	糖尿病

専門家と呼ばれる人の場合でも(1)

登録番号	組織型	悪性度	分化度	既往歴
1	10	0	1	0
2	11	1	2	0
3	10	1	9	0
4	21	2	1	0
5	22	1	2	1
6	11	1	1	0
7	10	2	99	0

専門家と呼ばれる人の場合でも(1)

登録番号	組織型	悪性度	分化度	既往歴
1	10	0	1	0
2	11	1	2	0
3	10	1	9	0
4	21	2	1	0
5	22	1	2	1
6	11	1	1	0
7	10	2	99	0

専門家と呼ばれる人の場合でも(2)

- 全角と半角
- 入力の仕方が統一されていない
- フィールド内で改行

登録番号	12	
転帰	1	死亡
最終生存確認日	2004/5/1	
死亡日	2004/6/10	
主死因	1	原病死
死因時の状況	本人の希望で自宅近くの病院へ転院、、、、○×○×○×。徐々に全身状態悪化し、摂食困難となり×× ×× ×× ×× ×× ×× ×× ×× ××。	

見えないけど、これ。
なぜ困る？

ここまでのまとめ

- いくつかの例より
 - 何かをするときに困る
 - 何が困る？
-
- その困ったことはなぜ起こる？

始めから考えておかないと後が大変

- ではどうすればいいか
 - データ構造を理解する
 - CRF \doteq DB を意識する

始めから考えておかないと後が大変

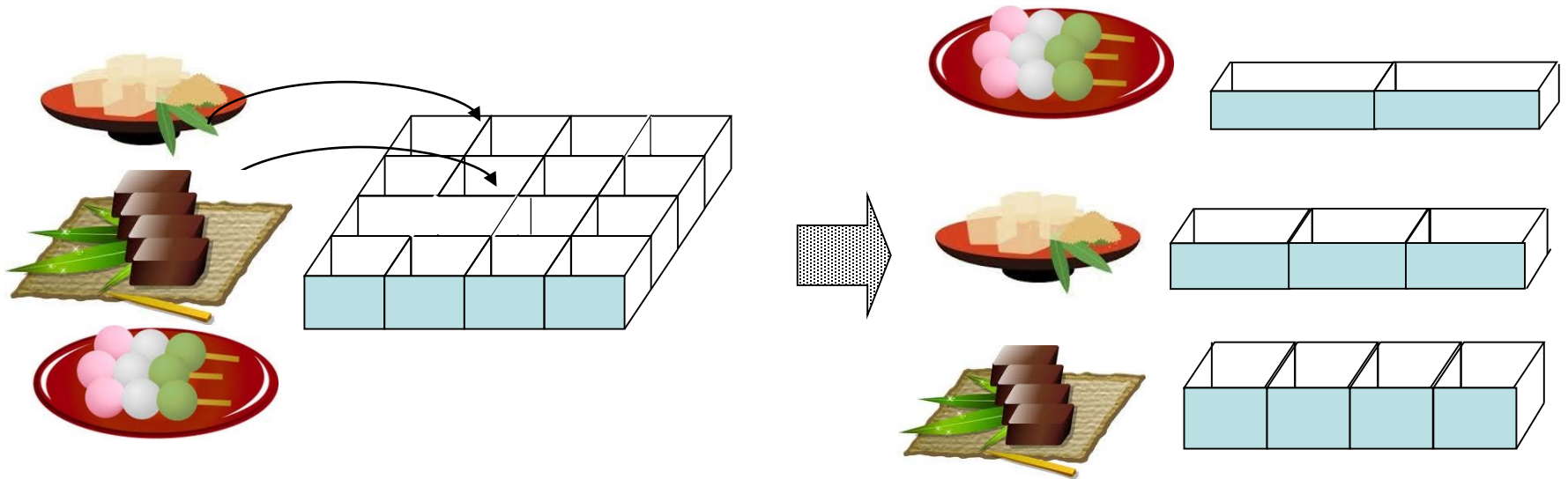
- 問題

- データ型が統一されていない
- 余計なテキストが入っている
- 全角と半角
- フィールド内で改行

- 解決策

- データ定義を知る

データ定義



- データに関する入れ物を定義する
- 各データの型や属性などを定義する

データ定義

入力画面

登録番号	<input type="text"/>
検査日	<input type="text"/>
白血球数	<input type="text"/>



テーブル

	CASENO	LABODATE	WBC
1	2015/7/1	4500	
2		2015/7/5	6400
3		2015/7/3	5500

No	フィールド名	ラベル	タイプ	サイズ	入力形式
1	CASENO	登録番号	テキスト型	32	テキストボックス
2	LABODATE	検査日	日付/時刻型	20	テキストボックス
3	WBC	白血球数	整数型	6	テキストボックス

レコード（行）を一意に識別するフィールド（キー）が必要

データ構造の種類

- 基本は、たった3つ！
 - ① Single タイプ
 - ② Visit タイプ
 - Visit – Item タイプ
 - ③ Event タイプ
- いずれも「キー」が必要
 - レコードを一意に識別するフィールド

Single タイプ

- 1 人に 1 回のみ観察されるもの
 - 例) 性別、生年月日、イニシャル など
 - 1 人 1 レコード

登録番号	イニシャル	性別	生年月日	疾患名
1	SH	男性	1970/2/5	喘息
2	YA	女性	1955/7/3	高血圧
3	KO	男性	1945/12/1	糖尿病

Visit タイプ

- 1 人に複数回観察され、複数のキーで管理
 - 例) 臨床検査値、有害事象
 - 1 人につき複数レコード(時系列の繰り返し)

登録番号	観察日	WBC	PLT	GPT
1	2004/8/1	4700	23.1	23
1	2004/10/1	6200	19.5	14
2	2005/5/20	8300	20.8	30
2	2005/6/29	5900	24.4	29

Visit-Item タイプ

- 1 人に複数回観察され、複数のキーで管理
 - 例) 臨床検査値、有害事象
 - 1 人につき複数レコード(項目の繰り返し)

登録番号	観察日	検査項目	結果
1	2004/8/1	WBC	4700
1	2004/8/1	PLT	23.1
1	2004/10/1	WBC	6200
1	2004/10/1	PLT	19.5
2	2005/5/20	WBC	8300
2	2005/5/20	PLT	20.8

Event タイプ

- 1 人に複数回観察され、観察内容がキーになるイベントが起こったときのみ記録
 - 例) 有害事象
 - 1 人につき複数レコード(項目の繰り返し)

症例番号	有害事象	Grade	発現日
1	アレルギー	3	2004/12/8
1	発熱	2	2004/12/14
3	食欲不振	2	2005/4/15

始めから考えておかないと後が大変

- 問題

- 識別子がない
- 入力の仕方が統一されていない
- 全角と半角

- 解決策

- どう集計・解析されるかを知る

識別子

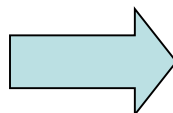
- 化学療法と放射線治療の毒性を分けて集計したい

	白:化学療法	青:放射線治療
登録番号	治療開始日	治療終了日
1	2009/2/14	2009/3/10
1	2009/3/23	2009/4/2
2	2009/3/31	2009/4/10
2	2009/4/21	2009/4/30
3	2009/4/1	2009/4/21
4	2009/4/10	2009/5/1

入力の仕方が統一されていない

テーブル

登録番号	性別
1	男
2	女
3	男性
4	女
5	女性

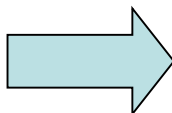


集計表

性別	人数

集計結果は・・・？

登録番号	性別
1	男
2	女
3	男性
4	女
5	女性



結果①

性別	人数
男性	2
女性	3

結果②

性別	人数
男	2
女	3

結果③

性別	人数
男	1
男性	1
女	2
女性	1

全角と半角

- 何が違う？
 - サイズが違う
 - 英数字、記号は1バイトで表現
 - コンピュータはすべてのデータを 0 と 1 で表現(2進数)
 - 2進数で何桁になっているかをビットで表す(8 ビット = 1 バイト)
 - 日本語は2バイトで表現
 - シフトJIS: Windows、Mac OS
 - 半角とは、バイト数が 1 の文字
 - 全角とは、バイト数が 2 の文字

半角	1	1 1 1 1 0 0 0 1
	A	1 1 0 0 0 0 0 1
全角	1	0 0 1 0 0 0 1 1 0 0 1 1 0 0 0 1
	A	0 0 1 0 0 0 1 1 0 1 0 0 0 0 0 1

フィールド内の改行

入力画面

中止判定日: 終了理由: 詳細:

1 プロトコル 本人の希望で自宅近くの病院へ転院。
ああああ

テーブル

症例番号	ページ番号	レポートID	レポート番号	登録番号	Visit番号	時期	中止判定日	食道切除有	食道切除が できない	終了理由	詳細
0001	73	1	1	1	8	alltxend		1		1	本人の希望で自宅近くの病院へ転院。 ああああ

テーブル内のデータを Excel ファイル(csv ファイル)にした場合

症例番号	ページ番号	レポートID	レポート番号	登録番号	Visit番号	時期	中止判定日	食道切除有	食道切除が できない	終了理由	詳細
1	63	1	1	1	6	neotxend				1	
1	73	1	1	1	8	alltxend			1	1	本人の希望で自宅近くの病院へ転院。 ああああ

解決のための手段

- データベース定義書(DB 定義書)
- Annotated CRF

データベース定義書

• DB 定義書の構造

TBLNAME	TBLTYPE	FLDNAME	FLDLABEL	FLDSIZE	FLDTYPE	CRF名	対応するCRFデータ項目
PRETREAT	1症例1レコード	TOBACCO	喫煙歴有無	5	整数型	治療前報告1	喫煙歴有無
PRETREAT	1症例1レコード	TOB_YRS	喫煙年数	5	整数型	治療前報告1	喫煙年数
PRETREAT	1症例1レコード	TOB_NUM	1日平均本数	5	整数型	治療前報告1	1日平均本数

コード名	コード	取り得るコード	入力ルール	例外入力	備考	履歴
有無	0:なし、1:あり	0、1				

データベース定義書

- 何に役立つ？
 - コミュニケーションツール
 - テーブル作成
 - 統計解析
 - 品質管理
 - ルールの記述
 - データクリーニング
 - 仕事の文書化のひとつの形

データクリーニングの例

FL DN AM E	FLDLABEL	FLDTYPE	FLDSI ZE	SL AB	PT YP	CRF SO RT	CRF名	対応するCRF データ項目	コード名	コード	取り得るコード	入力ルール
-	P_NU	テキスト型	6	-	-		手術・病理所見記録	N因子		00:pN0, 10:pN1, 99:pNX	00, 10, 99	
-	P_MU	テキスト型	6	-	-		手術・病理所見記録	M因子		00:pM0, 10:pM1, 11:pMA, 99:pMX	00, 10, 11, 99	
-	P_RU	整数型	5	-	-		手術・病理所見記録	R因子		0:R0, 1:R1, 2:R2, 99:RX	0, 1, 2, 99	
-	PGRWTH	整数型	5	-	-		手術・病理所見記録	発育様式		1:内向性, 2:外向性, 3:内向性 + 外向性, 4:内向性とも外向性とも判断できず, 5:頸管内病変のため見えず, 9:不明	1, 2, 3, 4, 5, 9	

	A	B	C	D	E	F	G	H
1	CASENC	OSXDT	SXMTD	PSTOPT	P_TU	P_NU	P_MU	P_RU
2	1	1998/04/21	1	0	121	00	昇順で並べ替え 降順で並べ替え	0
3	2	1998/01/09	4	0	210	00		0
4	3	1998/02/03	4	1	220	10	(すべて)	0
5	4	1998/03/06	4	0	121	00	(トップテン...)	0
6	5	1998/04/07	4	0	121	00	(オプション...)	0
7	6	1998/10/27	2	1	121	00	0	0
8	7	1998/11/10	2	0	121	99	000	0
9	8	1998/07/03	2	0	121	99	10	0
							11	
							99	0

コードとは

- 文字、数字、記号、およびそれらを関連づける規則の集合体。一連の文字列を識別、選択するために利用
- 少ない種類の記号で表現
- 短く表現
- 正確に表現
- 計算可能

－ 例) 男性: 1 男性: M
 女性: 2 女性: F

Annotated CRF

- データベース上の
テーブル名と変数名（フィールド名）や
その定義情報を CRF のレイアウト上に
記入したもの
- データベース定義書に比べて
 - わかりやすい
 - イメージしやすい（入力フォーム作成など）
 - 複数回収集するデータ
 - 時期、頻度を把握しやすい

Annotated CRFの例

JCOG 0910		大腸がん/ CRC Adj-CAPS		調査	
追跡調査用紙				までにデータセンターに郵送	
スクリーン名: FOLLOW UP スクリーンタイトル: 追跡調査 (カルテ番号)		がんセンター 患者イニシャル 姓 A 名 A		担当医 <input type="radio"/> X <input type="radio"/> X 登録番号 ****	
記入者: CRC記入可(医師)					
年 月 日					
[PROGRESS]					
再発の有無 <input type="checkbox"/> 再発なし または 不明		最終初再発生存確認日		NREC.DT 年 月 日	
(初回再発) <input type="checkbox"/> 再発あり REC		再発判定日		SEC.DT 年 月 日	
・2回目以降の再発については「後治療」の欄にご記入ください。					
再発部位		<input type="checkbox"/> 吻合部 Q_ANAST <input type="checkbox"/> 肝 Q_LIVER <input type="checkbox"/> 肺 Q_LUNG <input type="checkbox"/> 肝・肺以外の血行 Q_HEMAT <input type="checkbox"/> 腹膜 Q_PERIT <input type="checkbox"/> 遠隔リンパ節内 Q_REGLN <input type="checkbox"/> リンパ管性(遠隔リンパ節以外のリンパ管) Q_OTHLN <input type="checkbox"/> その他 Q_OTH <input type="checkbox"/> 不明 Q_UNKNOWN			
前診調査の報告		状況 Q.MEMO			
結膜内癌 または Carcinoma in situ(上皮内癌)の有無 REC1 <input type="checkbox"/> なし <input type="checkbox"/> 初回の病変あり 最終初病変生存確認日 NSEC.DT1 年 月 日 初回の病変判定日 REC.DT1 年 月 日 病変部位					
二次がん/同時性重複癌 結膜内癌・Carcinoma in situ以外のがん病変の有無 SEC <input type="checkbox"/> がん病変なし <input type="checkbox"/> がん病変あり 最終初病変生存確認日 NSEC.DT 年 月 日 がん病変判定日 SEC.DT 年 月 日 病変部位 S SITE					
前診調査までの報告		状況 Q.MEMO (共有)			
[AFTREAT] CASENO PERIOD=YYYYMMPP					
後治療 <input type="checkbox"/> なし <input type="checkbox"/> 後治療あり <input type="checkbox"/> 不明		後治療開始日		AF.STDT 年 月 日	
AFTREAT		<input type="checkbox"/> 化学療法 AF_CX <input type="checkbox"/> 外科的手術 AF_SX <input type="checkbox"/> その他 AF_OTH		内容など詳細 AF.TREAT	
[DTINFO] CASENO					
転移		前診調査の報告			
SURVIVE <input type="checkbox"/> 生存 <input type="checkbox"/> 死亡		最終生存確認日		FINALDT 年 月 日	
死亡日		DEATHDT 年 月 日			
D_CAUSEL 死因 <input type="checkbox"/> 原病死 <input type="checkbox"/> 他病死 <input type="checkbox"/> 治療関連死 <input type="checkbox"/> 後治療による治療関連死 <input type="checkbox"/> その他 <input type="checkbox"/> 不明					
死亡の状況		MEMO			
いずれの死因の報告も死亡時の状況を記入					

まとめ

- データベースは、ある規則に従ってまとめられたデータの集まり
- データベース作成のためにデータを定義
 - データベース定義書
 - Annotated CRF
- 後々の解析がスムーズに進められる